Claims

We claim:

1. A method for segmenting a video including a plurality of pixels into a plurality of video objects, comprising:

assigning a feature vector to each pixel of the video;

identifying selected pixels of the video as marker pixels;

assembling each marker pixel and pixels adjacent to the marker pixel into a corresponding a volume if the distance between the feature vector of the marker pixel and the feature vector of the adjacent pixels is less than a first predetermined threshold;

assigning a first score and descriptors to each volume;

sorting the volumes in a high-to-low order according to the first scores; and

processing the volumes in the high-to-low order, the processing for each volume comprising:

comparing the descriptor of the volume to the descriptor of an adjacent volume to determine a second score;

combining the volume with the adjacent volume if the second score passes a second threshold to generate a video object in a multi-resolution video object tree; and

repeating the comparing and combining steps until a single video representing the video remains.

2. The method of claim 1 wherein each pixel has spatial (x,y) and time (t) coordinates to indicate a location of the pixel and the volumes in a spatial-temporal collocated overlapping scene of the video.

3. The method of claim 2 wherein the video includes a plurality of frames and further comprising:

projecting a portion of each video object in a particular frame to intersect the projection of the video object in an adjacent frame to provide continuous silhouettes of the video object according to the time t coordinates.

4. The method of 3 further comprising:

applying a spatial-domain 2D median filter 210 to the frames 102 to remove intensity singularities, without disturbing edge formation.

5. The method of claim 1 further comprising:

partitioning the video into a plurality of identically sized volumes; and

selecting the pixel at the center of each volume are the marker pixels.

6. The method of claim 1 further comprising:

determining a gradient magnitude $\nabla V = \partial V/\partial x + \partial V/\partial y + \partial V/\partial t$ for each pixel in the video;

selecting the pixel with a minimum gradient magnitude as the marker pixel;

removing pixel in a predetermined neighborhood around the marker; and

repeating the selecting and removing steps until no pixel remain.

7. The method of claim of claim 1 wherein the feature vector is based on a color of the pixel.

8. The method of claim 1 further comprising:

merging volumes less than minimum size with an adjacent volumes.

9. The method of claim 8 wherein the minimum size is less than 0.001 of the volume representing the video.

10. The method of claim 9 further comprising:

sorting the volumes in an increasing order to size;

processing the volumes in the increasing order, the processing for each volume comprising:

including each pixel of the volume less in a closest volume until all volumes less than the minimum size are processed.

11. The method of claim 1 wherein the descriptors include self descriptors of the volume, and mutual descriptors of the volume and the adjacent volume.

12. A method for segmenting a video sequence of frames, each frame including a plurality of pixels, comprising:

partitioning all of the pixels of all frames of the video into a plurality of volumes according to features of each pixel, the pixels of each volume having frame-based spatial coordinates and sequence-based temporal coordinates;

assigning descriptors to each volume;

representing each volume as a video object at a lowest level in a multi-resolution video object tree; and

iteratively combining volumes according to the descriptors, and representing each combined volume as a video object at intermediate levels of the multi-resolution video object tree, until all of the combined volumes form the entire video represented as a video object at a highest level of the multi-resolution video object tree.